Electronic Colloquium on Computational Complexity
ECCC Technical Reports Series 1994 Report Nr: TR94-001

On Rank vs. Communication Complexity[†]

Noam Nisan [‡] Avi Wigderson [§]

Received November 1, 1994

Abstract. This paper concerns the open problem of Lovasz and Saks regarding the relationship between the communication complexity of a boolean function and the rank of the associated matrix. We first give an example exhibiting the largest gap known. We then prove two related theorems.

Keywords: Communication Complexity

[†] Extended Abstract appeared in FOCS 1994.

[‡] Institute of Computer Science, Hebrew University of Jerusalem, Israel. This work was supported by USA-Israel BSF grant 92-00043 and by a Wolfeson research award administered by the Israeli Academy of Sciences.

[§] Institute of Computer Science, Hebrew University of Jerusalem, Israel. This work was supported by USA-Israel BSF grant 92-00106 and by a Wolfeson research award administered by the Israeli Academy of Sciences.

1 Introduction

For a $0,1$ matrix M, denote by $c(M)$ the deterministic communication complexity of the associated function [Y79], and by $rk(M)$ its rank over the reals. It is well known [MS82] that $\log rk(M) \leq c(M) \leq rk(M)$. It is a fundamental question of communication complexity to narrow this exponential gap. As rank arguments are the main source of deterministic communication complexity lower bounds, and the rank function has many useful properties, it would make life nicer if the lower bound was rather tight. A tempting conjecture (see [LS88]) is

Conjecture 1 *For every matrix M, $c(M) = (\log rk(M))^{O(1)}$*

Lovász and Saks [LS89] also show that this conjecture is strongly related to a conjecture of van Nuffelen [Nu76] and Fajtlowicz [Fa87] regarding the connection between the chromatic number of a graph and the rank of its adjacency matrix.

Several authors have obtained separation results between $c(M)$ and $\log rk(M)$ [AS89, Raz92]. The best separation known so far gives an infinite family of matrices for which $c(M) \geq \log rk(M) \log\log\log rk(M)$ [RS93]. Our first result is an example with a much larger gap.

Theorem 1 *There exist (explicitly given) 0-1 matrices M of size $2^n \times 2^n$ such that $c(M) = \Omega(n)$, and $\log rk(M) = O(n^\alpha)$, where $\alpha = \log_3 2 = 0.63...$*

The same $\Omega(n)$ lower bound applies also to the randomized and to the nondeterministic communication complexities. The construction is based on boolean functions with high "sensitivity" and low degree. Such a function was constructed in [NS92]. The lower bound for the communication complexity relies on the known lower bounds for randomized communication complexity of "disjointness" [KS87, Raz90]. Recently Kushilevitz [Ku94] has somewhat improved the construction of [NS92] and has thus reduced the value of α to $\log_6 3 = 0.61...$. The main lemma of [NS92] shows however that this technique cannot reduce the value of α to below $1/2$.

We then return our attention to conjecture 1, and consider weaker related conjectures. To explain them, we need some notation. If S is a subset of the entries of M, let S_0 and S_1 denote respectively the subsets of S whose value is 0 and 1 respectively. Call S *monochromatic* if either $S = S_0$ or $S = S_1$.

Let $mono(M)$ denote the maximum fraction $|A|/|M|$ over all monochromatic submatrices A of M. When S is not monochromatic, we will be interested in the advantage one color has over the other. The *(absolute) discrepancy* of S is $\delta(S) = |(|S_0| - |S_1|)/|M||$. Define $disc(M)$ to be the maximum of $\delta(A)$ over all submatrices A of M.

Since an optimal protocol for M partitions it into at most $2^{c(M)}$ monochromatic rectangles, we have the basic relation:

$$disc(M) \geq mono(M) \geq 2^{-c(M)}$$

or, equivalently,

$$-\log disc(M) \leq -\log mono(M) \leq c(M).$$

Thus two conjectures weaker than Conjecture 1 suggest themselves. They respectively assert that low rank matrices have large monochromatic rectangles, or weaker still, large discrepancy.

Conjecture 2 *For every M,* $-\log mono(M) = (\log rk(M))^{O(1)}$

Conjecture 3 *For every M,* $-\log disc(M) = (\log rk(M))^{O(1)}$

As mentioned, Conjecture 1 → Conjecture 2 → Conjecture 3. We first prove, in theorem 2, that conjectures 1 and 2 are equivalent. We then prove, in theorem 3, (a strong form of) conjecture 3.

Theorem 2 *Conjecture 1* iff *Conjecture 2.*

Thus in order to prove conjecture 1 it suffices to show that every low rank boolean matrix has a "large" monochromatic submatrix. In fact, the proof of the theorem implies that it suffices to show that every rank r boolean matrix has a "large" submatrix of rank at most, say, $0.99r$.

Theorem 3 *For every M,* $1/disc(M) = O(rk(M)^{3/2})$.

Note that Theorem 3 implies Conjecture 3. The bound in this theorem is nearly tight: for every r there are infinitely many matrices M of rank r and $1/disc(M) \geq r$. This can be easily seen by taking any square array of $r \times r$ Hadamard matrices.

This theorem supplies the first clue that low rank has something to do with low communication complexity, though in a very weak sense. The communication model we have in mind is distributional communication complexity, where the inputs are chosen at random [Y83]. For this model, low rank guarantees a cheap protocol with a nontrivial advantage over guessing the function value. In the protocol each player sends one bit specifying whether or not his input is in the biased rectangle. Precisely:

Corollary 1 *If $rk(M) = r$, then there is a 2 bit protocol P, which satisfies $Pr[P(x,y) = M(x,y)] \geq 1/2 + \Omega(1/r^{3/2})$, where the input (x,y) is chosen uniformly at random.*

2 Proof of Theorem 1

We will require the following definition.

Defininition: Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function. We say that f is fully sensitive at $\vec{0}$ if $f(\vec{0}) = 0$ and yet for any vector x of hamming weight 1 (i.e. for any unit vector), $f(x) = 1$.

The degree of f, $deg(f)$ is defined to be the degree of the unique multivariate multi-linear polynomial over the reals which agrees with f on $\{0,1\}^n$.

In [NS92] it is shown that any boolean function which is fully sensitive at $\vec{0}$ must have degree of at least $\sqrt{n}/2$. They also give an example of a fully sensitive function with degree significantly less than n.

Lemma 1 *[NS92] There exists an (explicitly given) boolean function $f : \{0,1\}^n \to \{0,1\}$ which is fully sensitive at $\vec{0}$ and $deg(f) = n^\alpha$, for $\alpha = \log_3 2 = 0.63\ldots$. Furthermore, f has at most $2^{O(n^\alpha)}$ monomials.*

For completeness we repeat the construction of [NS92].

Proof: Let $E(z_1, z_2, z_3)$ be the symmetric boolean function giving 1 iff exactly 1 or 2 of its inputs are 1. It is easy to check that E is fully sensitive at $\vec{0}$. One may also readily verify that $deg(E) = 2$ as $E(z_1, z_2, z_3) = z_1 + z_2 + z_3 - z_1z_2 - z_1z_3 - z_2z_3$. We now recursively define a function E_k

on 3^k input bits by: $E^0(z) = z$, and $E^k(\cdot) = E(E^{k-1}(\cdot), E^{k-1}(\cdot), E^{k-1}(\cdot))$, where each instance of E^{k-1} is on a different set of 3^{k-1} input bits. It is easy to prove by induction that (1) E^k is fully sensitive at $\vec{0}$, (2) $deg(E^k) = 2^k$, and (3) E^k has at most 6^{2^k-1} monomials. Our desired f is the function E^k on $n = 3^k$ variables[1]. □

We now transform f into a matrix as follows.

Definition: With every boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$ we associate a $2^n \times 2^n$ matrix M_f as follows:

$$M_f(x_1 \dots x_n; y_1 \dots y_n) = f(x_1 \cdot y_1, x_2 \cdot y_2 \dots x_n \cdot y_n)$$

The properties of M_f are ensured by the following lemmas.

Lemma 2 *If f is fully sensitive at $\vec{0}$ then $c(M_f) = \Omega(n)$. The same lower bound holds for the randomized and for the nondeterministic complexity of M_f.*

Lemma 3 *Let f be a polynomial with m monomials, then $rk(M_f) \leq m$. In particular, if $d = deg(f)$ then $rk(M_f) \leq \sum_{i=0}^{d} \binom{n}{i} = 2^{O(d \log n)}$.*

Proof (of lemma 2): This proof is a direct reduction from the known lower bounds for the randomized communication complexity of disjointness. These bounds actually show that it is even hard to distinguish between the case where the sets are disjoint and the case where the intersection size is 1.

Let the $UDISJ$ problem be the following: the two players are each given a subset of $\{1 \dots n\}$. If the sets are disjoint they must accept. If the sets intersect at exactly 1 point then they must reject. If the size of the intersection is greater than 1 then the players are allowed to either accept or reject.

Theorem ([KS87], see also [Raz90]): Any communication complexity protocol for $UDISJ$ requires $\Omega(n)$ bits of communication. The same is true for non-deterministic and for randomized protocols.

[1] Recently, [Ku94] has improved upon this construction by exhibiting a function E' on 6 variables which is fully sensitive at $\vec{0}$ and with degree only 3. Using the same recursion, this reduces α to $\log_6 3 = 0.61...$

Now notice that if f is fully sensitive at $\vec{0}$ then any protocol for M_f directly solves $UDISJ$. This is done by transforming each set to its characteristic vector. If the sets are disjoint then for each i, $x_i y_i = 0$, and thus $M_f(\vec{x}, \vec{y}) = f(\vec{0}) = 0$. If the intersection size is exactly 1 then in exactly 1 position $x_i y_i = 1$, and thus $M_f(\vec{x}, \vec{y}) = 1$. □

Proof (of lemma 3): Let $f(z_1 \ldots z_n) = \sum_S \alpha_S \prod_{i \in S} z_i$ be the representation of f as a real polynomial. By the definition of M_f it follows that $M_f = \sum_S \alpha_S M_S$, where the matrix M_S is defined by $M_S(\vec{x}, \vec{y}) = \prod_{i \in S} x_i \cdot y_i$. But clearly for each S, $rk(M_S) = 1$. It follows that the rank of M_f is bounded from above by the number of non-zero monomials of f. The bound in terms of the degree follows directly. □

The combination of lemmas 2 and 3 with the function E^k constructed in lemma 1 gives the statement of the theorem. □

3 Proof of Theorem 2

Assume conjecture 2, i.e. assume that every $0,1$ matrix M has a monochromatic submatrix of size $|M|/exp(\log^k rk(M))$. Given a $0,1$ matrix M we will design a communication protocol for M.

Let A be the largest monochromatic submatrix of M. Then A induces in a natural way a partition of M into 4 submatrices A, B, C, D, with B sharing the rows of A and C sharing the columns of A. Clearly $rk(B) + rk(C) \leq rk(M) + 1$. Assume w.l.o.g. that $rk(B) \leq rk(C)$, then the submatrix $(A|B)$ has rank at most $2 + rk(M)/2$.

In our protocol the row player sends a bit saying if his input belongs to the rows of A or not. The players then continue recursively with a protocol for the submatrix $(A|B)$, or for the submatrix $(C|D)$, according to the bit communicated.

Denote by $L(m, r)$ the number of leaves of this protocol, starting with a matrix of area at most m and rank at most r. By the protocol presented we get a recurrence $L(m, r) \leq L(m, 2 + r/2) + L(m(1 - \alpha), r)$, where α is the fraction of rows in A. By the assumption, $\alpha \geq (exp(\log^k r))^{-1}$. Note that (assuming the players ignore identical rows and columns) that $m \leq 2^r$, and that $L(m, 1) = 1$. It is standard to see that the solution to the recurrence satisfies $L(m, r) \leq exp(\log^{k+1} r)$.

We have so far obtained a protocol for M with $exp(\log^{k+1} rk(M))$ leaves; it is well known that this implies also $c(M) \leq O(\log^{k+1} rk(M))$. □

Remark: Note that the same proof, yielding essentially the same bound, would go through even if instead of a large monochromatic (rank 1) submatrix we were promised a large submatrix of rank $r/4$, say. The idea is that for the decomposition A, B, C, D in the proof we have in general $rk(B) + rk(C) \leq rk(M) + rk(A)$. We used it above for a monochromatic A, so $rk(A) \leq 1$. Now we have $rk(A) \leq r/4$, and using $rk(B) \leq rk(C)$ we get $rk(B) \leq (rk(M) + rk(A))/2 \leq 5r/8$. Thus $rk(A|B) \leq rk(A) + rk(B) \leq 7r/8$. The recurrence relation changes to $L(m, r) \leq L(m, 7r/8) + L(m(1-\alpha), r)$, which has the same asymptotic behavior.

The expression $r/4$ may be raplaced by αr for any $\alpha < 1$ by repeatedly taking a large submatrix of low rank of the current submatrix. After constant number of times the rank is reduced to $r/4$. Again, this does not change the asymptotics of the recurrence.

4 Proof of Theorem 3

Let us consider $-1, +1$ matrices rather than $0, 1$ matrices; this obviously changes the rank by at most 1, and does not change the discrepancy. The advantage is that the discrepancy of a submatrix N of M has a simple form: $\delta(N)$ is the sum of entries of N, divided by the area of M.

We will use the following notation. Let $x = (x_i) \in R^n$ and $A = (a_{ij})$ be an $n \times n$ real matrix. Then:

- $||x|| = (\sum_{i=1}^n x_i^2)^{1/2}$, the L_2 norm of x.
- $||x||_\infty = max_{i=1}^n |x_i|$, the L_∞ norm of x.
- $||A|| = max_{||x||=1} ||Ax||$, the spectral norm of A. It is well known that also $||A|| = max_{||x||=1, ||y||=1} |x^T Ay|$; and $||A|| = max\{\sqrt{\lambda} : \lambda \text{is an eigenvalue of } A^T A\}$.
- $W(A) = (\sum_{i,j=1}^n a_{ij}^2)^{1/2}$, the Euclidean norm of A.
- $tr(A) = \sum_{i=1}^n a_{ii}$, the trace of A.

Overview of Proof: It is best to summerize the proof backwards. We are given a ± 1 matrix A of low rank and wish to find in it a submatrix of high discrepancy. This is done in lemma 6 and is clearly equivalent to finding $0,1$ vectors x and y such that $x^T Ay$ is large. As an intermediate step we shall, in lemma 5, find real vectors u and v, having low L_∞-norm, with $u^T Av$ large. Towards this we shall need real vectors w and z having low L_2-norm, with $w^T Az$ large. This is equivalent to proving lower bounds on $\|A\|$, which we do in lemma 4.

Lemma 4 *For every real matrix A,*

$$\frac{W(A)}{\sqrt{rk(A)}} \leq \|A\| \leq W(A)$$

Proof: Let $r = rk(A)$. Let us compute the trace of $A^T A$. On one hand, direct calculation by definition shows that $tr(A^T A) = W(A)^2$. On the other hand $tr(A^T A) = \sum_i \lambda_i$, where the sum is over all eigenvalues λ_i of $A^T A$. Since $A^T A$ has only r non-zero eigenvalues, and since all eigenvalues of $A^T A$ are positive, the largest eigenvalue, λ_1, is bounded by $\frac{W(A)^2}{r} \leq \lambda_1 \leq W(A)^2$. The lemma follows since $\|A\| = \sqrt{\lambda_1}$. □

Lemma 5 *Let A be an $n \times n$ ± 1 matrix of rank r. Then there exist vectors u, v, $||u||_\infty \leq 1$, $||v||_\infty \leq 1$, such that $u^T Av \geq \frac{n^2}{16r^{3/2}}$.*

Proof: Denote $r = rk(A)$. Let x and y be vectors such that $||x|| = 1$, $||y|| = 1$, and $x^T Ay = \|A\|$. Let $I = \{i : |x_i| > \sqrt{8r/n}\}$ and $J = \{j : |y_j| > \sqrt{8r/n}\}$. Notice that $|I| \leq n/(8r)$, and $|J| \leq n/(8r)$.

Let $\hat{u}$ be the vector that agrees with x outside of I and is 0 for indices in I, and let $\hat{v}$ be the vector that agrees with y outside of J and is 0 for indices in J.

We shall compute a lower bound on $\hat{u}^T A\hat{v}$. Consider the matrix B defined to agree with A on all entries i, j such that $i \in I$ or $j \in J$, and to be 0 elsewhere. Using this notation it is clear that

$$\hat{u}^T A\hat{v} = x^T Ay - x^T By.$$

A lower bound for $x^T Ay = \|A\|$ is obtained using the lower bound in lemma 4, and as $W(A) = n$, $x^T Ay \geq n/\sqrt{r}$. An upper bound for $x^T By$ is given by the upper bound in the last lemma $x^TBy \leq \|B\| \leq W(B)$. Since B has at most $n/(8r)$ non-zero rows and $n/(8r)$ non-zero columns, $W(B) \leq n/(2\sqrt{r})$. It follows that $\hat{u}^TA\hat{v} \geq n/(2\sqrt{r})$.

Now define $u = \sqrt{n/(8r)}\hat{u}$ and $v = \sqrt{n/(8r)}\hat{v}$. By definition $||v||_\infty \leq 1$ and $||u||_\infty \leq 1$. The lemma follows since $u^TAv = n/(8r)\hat{u}^TA\hat{v}$. □

Lemma 6 *Let A be an $n \times n$ matrix, and u, v vectors such that $||u||_\infty \leq 1$, $||v||_\infty \leq 1$. Then there exists a submatrix B of A with $\delta(B) \geq u^TAv/(4n^2)$.*

Proof: Let $z = Av$. Clearly, $\sum_{i \in K} u_i z_i \geq u^TAv/2$, where K is either the coordinates where both u_i and z_i are positive or the coordinates in which both are negative. Assume the first case (otherwise replace below $v \leftarrow -v$). Then setting $x = \chi_K$ (the characteristic vector of K), we have (using $||u||_\infty \leq 1$), $x^TAv \geq u^TAv/2$. Repeating this argument with $z = x^TA$, we can replace v with a $0, 1$ vector y obtaining $x^TAy \geq u^TAv/4$. Now take B to be the submatrix defined by the 1's in x and y. Since B is a ±1 matrix, the bilinear form divided by n^2 gives its discrepancy. □

Combining lemmas 5 and 6, every ±1 matrix A of rank r, contains a submatrix B with $\delta(B) \geq \frac{1}{64r^{3/2}}$. Thus $disc(M) \geq \frac{1}{64r^{3/2}}$, and theorem 3 follows. □

Acknowledgements

We thank Oded Goldreich, Russell Impagliazzo, Mauricio Karchmer, Eyal Kushilevitz, and Roman Smolensky for conversations on this subject.

References

[AS89] N. Alon, P. Seymour, "A counter example to the rank-covering conjecture", J. Graph Theory 13, pp. 523–525, 1989.

[Fa87] S. Fajtlowicz, "On conjectures of Graffiti" II, *Congresus Numeratum* 60, pp. 189–198, (1987).

[KS87] B. Kalyanasundaram and G. Schnitger, "The probabilistic communication complexity of set intersection", 2*nd Structure in Complexity Theory Conference*, pages 41–49, 1987.

[Ku94] E. Kushilevitz, manuscript, 1994.

[LS88] L. Lovász and M. Saks, "Lattices, Möbius functions, and communication complexity", *Proc. of the 29th FOCS*, pp. 81–90, 1988.

[LS89] L. Lovász and M. Saks, Private communication.

[MS82] K. Mehlhorn, E.M. Schmidt, "Las Vegas is better than determinism in VLSI and distributive computing", *Proceedings of* 14th *STOC*, pp. 330-337, 1982.

[NS92] N. Nisan and M. Szegedy, "On the degree of boolean functions as real polynomials", *Proceedings of* 24th *STOC*, pp. 462–467, 1992.

[Nu76] C. van Nuffelen, "A bound for the chromatic number of a graph", *American Mathematical Monthly* 83, pp. 265–266, 1976.

[Raz90] A. Razborov, "On the distributional complexity of disjointness", *Proceedings of the ICALP*, pp. 249–253, 1990. (to appear in *Theoretical Computer Science*).

[Raz92] A. Razborov, "The gap between the chromatic number of a graph and the rank of its adjacency matrix is superlinear", Discrete Math. 108, pp 393–396, 1992.

[RS93] R. Raz and B. Spiker, "On the Log-Rank conjecture in communication complexity", *Proc. of the 34th FOCS*, pp. 168–176, 1993

[Y79] A. C.-C. Yao, "Some complexity questions related to distributive computing", *Proceedings of* 11th *STOC*, pp. 209-213 (1979).

[Y83] A. C.-C. Yao, "Lower Bounds by Probabilistic Arguments", *Proc. 24th FOCS*, pp. 420–428, (1983).